



Известия высших учебных заведений. Прикладная нелинейная динамика. 2021. Т. 29, № 5
Izvestiya Vysshikh Uchebnykh Zavedeniy. Applied Nonlinear Dynamics. 2021;29(5)

Научная статья
УДК 530.182
DOI: 10.18500/0869-6632-2021-29-5-799-811

Динамика искусственной рекуррентной нейронной сети в задаче моделирования когнитивной функции

О. В. Масленников

Институт прикладной физики РАН, Нижний Новгород, Россия
E-mail: olmaov@ipfran.ru

Поступила в редакцию 26.02.2021, принята к публикации 11.05.2021,
опубликована 30.09.2021

Аннотация. Цель работы состоит в построении искусственной рекуррентной нейронной сети, активность которой моделирует когнитивную функцию, связанную со сравнением двух вибротактильных стимулов, предъявляемых с задержкой, и анализе динамических механизмов, лежащих в основе её работы. **Методы.** Машинное обучение, анализ пространственно-временной динамики и фазового пространства. **Результаты.** Активность обученной рекуррентной нейронной сети моделирует когнитивную функцию сравнения двух стимулов с задержкой. Модельные нейроны демонстрируют смешанную селективность в процессе выполнения задачи. В многомерной активности выделены компоненты, каждая из которых зависит от одного параметра задачи. **Заключение.** Настройка искусственной нейронной сети выполнять функцию, аналогичную экспериментально наблюдаемому процессу, сопровождается появлением динамических свойств модельных нейронов, аналогичных найденным в эксперименте.

Ключевые слова: рекуррентная нейронная сеть, машинное обучение, когнитивная нейронаука.

Благодарности. Построение модели выполнено в рамках Программы развития регионального научно-образовательного математического центра «Математика технологий будущего», проект 075-02-2020-1483/1. Анализ динамики выполнен при поддержке РФФИ (грант № 19-72-00112).

Для цитирования: Масленников О. В. Динамика искусственной рекуррентной нейронной сети в задаче моделирования когнитивной функции // Известия вузов. ПНД. 2021. Т. 29, № 5. С. 799–811.
DOI: 10.18500/0869-6632-2021-29-5-799-811

Статья опубликована на условиях Creative Commons Attribution License (CC-BY 4.0).

Dynamics of an artificial recurrent neural network for the problem of modeling a cognitive function

O. V. Maslennikov

Institute of Applied Physics
of the Russian Academy of Sciences, Nizhny Novgorod, Russia
E-mail: olmaov@ipfran.ru

Received 26.02.2021, accepted 11.05.2021, published 30.09.2021

Abstract. The purpose of this work is to build an artificial recurrent neural network whose activity models a cognitive function relating to the comparison of two vibrotactile stimuli coming with a delay and to analyze dynamic mechanisms underlying its work. *Methods* of the work are machine learning, analysis of spatiotemporal dynamics and phase space. *Results.* Activity of the trained recurrent neural network models a cognitive function of the comparison of two stimuli with a delay. Model neurons exhibit mixed selectivity during the course of the task. In the multidimensional activity, the components are found each of which depends on a certain task parameter. *Conclusion.* The training of the artificial neural network to perform the function analogous to the experimentally observed process is accompanied by the emergence of dynamic properties of model neurons which are similar to those found in the experiment.

Ключевые слова: recurrent neural network, machine learning, cognitive neuroscience.

Acknowledgements. Design of the model was carried out within the framework of the Program for the Development of the Regional Scientific and Educational Mathematical Center “Mathematics of Future Technologies”, project 075-02-2020-1483/1. Analysis of the dynamics was supported by the Russian Science Foundation (grant No 19-72-00112).

For citation: Maslennikov O.V. Dynamics of an artificial recurrent neural network for the problem of modeling a cognitive function. *Izvestiya VUZ. Applied Nonlinear Dynamics.* 2021;29(5):799–811. DOI: 10.18500/0869-6632-2021-29-5-799-811

This is an open access article distributed under the terms of Creative Commons Attribution License (CC-BY 4.0).

Введение

Математическое моделирование активности нейронных популяций мозга как предмет изучения вычислительной нейронауки интенсивно развивается вот уже более полувека [1, 2]. Построение моделей проводится на различных уровнях – от отдельных нейронов до крупномасштабных сетей – и с различной степенью детализации – начиная с морфологически правдоподобных пространственно протяженных нервных клеток до сетей достаточно абстрактных активных единиц. Выбор уровня построения модели зависит от природы и масштаба явления, которое требуется описать. В области исследования динамических механизмов нейронной активности, лежащих в основе выполнения когнитивных задач, в настоящее время наблюдается формирование парадигмы функциональных искусственных нейронных сетей, построенных с применением машинного обучения [3–8]. Суть её состоит в следующем. Выполняемая в ходе эксперимента испытуемым животным когнитивная задача формализуется в терминах целевой функции, преобразующей входные стимулы в требуемые выходные отклики. Затем задается искусственная нейронная сеть, состоящая из простых элементов, структура которой может выбираться с учетом анатомических особенностей. После чего происходит контролируемое обучение (с учителем), то есть настройка параметров этой сети выполнять целевое преобразование входов в выходы на основе методов глубокого обучения [9, 10]. Полученная сеть представляет собой многомерную динамическую систему, или конкретнее, сложную сеть простых нелинейных элементов – искусственных нейронов. Результирующую систему можно исследовать методами нелинейной динамики и теории сетей [11–14]. В частности, можно выявить в многомерном фазовом пространстве особые траектории, которые отвечают выполнению той или иной функции, можно анализировать распределение весов, кластеризацию связей, наличие модулей в структуре обученной сети [15].

Кроме того, полученную искусственную нейронную сеть можно анализировать методами, имеющими своё основание в экспериментальных исследованиях нейронных популяций мозга. Можно находить каузальные последствия удаления отдельных нейронов и связей, выявлять более или менее важные нейроны и связи с точки зрения выполнения сетью целевой функции и т. д.

В настоящей статье описанный подход раскрывается на примере конкретного эксперимента, являющегося классическим в области когнитивной нейронауки. Эксперимент связан с изучением свойств рабочей памяти обезьяны при выполнении задачи вибротактильной дискриминации [16]. В первом разделе дается краткое описание эксперимента. Во втором разделе описывается модель рекуррентной нейронной сети и в терминах модели формулируется изучаемая когнитивная задача. В третьем разделе анализируется нелинейная динамика обученной рекуррентной нейронной сети при выполнении целевой задачи. В Заключении обсуждаются основные выводы.

1. Когнитивная задача: сравнение вибротактильных стимулов

Рассмотрим основные моменты эксперимента, в котором анализируется активность нейронов головного мозга обезьян при выполнении ими задачи вибротактильной дискриминации [16]. Схема эксперимента приведена на рис. 1, *a*. Обезьяна концом пальца зафиксированной лапы касается металлического стержня, а свободной лапой держится за ручку (слева на рис. 1, *a* и соответствующая точка А на оси времени рис. 1, *b*). Далее со стержня она получает вибротактильный стимул частотой f_1 длительностью 0.5 с, затем несколько секунд (в среднем 3 с) продолжается период задержки, после которого вновь предъявляется стимул с такой же продолжительностью на частоте f_2 . Далее животное должно определить, является ли частота f_2 выше или ниже, чем f_1 , для чего свободную лапу отцепляет от ручки (точка В на рис. 1, *b*) и нажимает соответствующую кнопку (справа на рис. 1, *a* и соответствующая точка С на рис. 1, *b*). Стимулы представляют собой гармонические вибрации частотой 10...34 Гц.

В процессе эксперимента происходит регистрация электрической активности отдельных нейронов префронтальной коры головного мозга, что позволяет сопоставить динамику нейронов высшего отдела центральной нервной системы с различными стадиями выполняемой задачи. С точки зрения проводимых мозгом когнитивных операций, описанная задача включает в себя следующую цепочку действий: кодирование частоты первого стимула, поддержание его в рабочей памяти во время задержки, кодирование частоты второго стимула, сравнение её со следом в памяти, оставленным первым стимулом, передача результата сравнения на двигательную систему. Иногда эту задачу классифицируют как параметрическую рабочую память, в том смысле, что осуществляется запоминание вещественного параметра – частоты получаемого посредством осязания стимула (скалярное аналоговое значение), а после временной задержки происходит сравнение запомненного значения со второй скалярной аналоговой величиной.

В ходе экспериментальных работ установлено, что в числе нескольких сотен нейронов префронтальной коры, активность которых регистрируется, существуют группы, которые

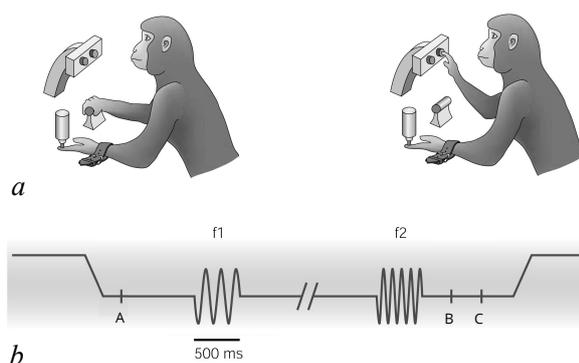


Рис. 1. *a* – Схема эксперимента с обезьяной, выполняющей задачу вибротактильной дискриминации. *b* – Последовательность событий в процессе эксперимента. Рисунок взят с изменениями из работы [16]

Fig. 1. *a* – Scheme of the experiment with a monkey performing the task of vibrotactile discrimination. *b* – The sequence of events during the experiment. The figure is modified from ref. [16]

ведут себя по-разному. Например, обнаружены нейроны, средняя частота спайковых разрядов которых является монотонной функцией частоты первого стимула, причем такая зависимость сохраняется во время интервала задержки. Есть группа нейронов, активность которых почти прекращается после предъявления первого стимула, а есть другая группа, нейроны которой в отсутствие стимулов продолжают генерировать спайки. Более того, часть из них достигает пика активности в момент действия первого стимула, после чего активность спадает, а другая часть демонстрирует активность, частота которой растет вплоть до предъявления второго стимула. Также есть нейроны с практически нулевой активностью вплоть до второго стимула, которые начинают генерировать выраженные разряды в момент принятия решения [16, 18].

Суммируя, можно сделать вывод, что в популяции нейронов префронтальной коры есть нейроны, которые по-разному отслеживают различные стадии выполняемой задачи, и их активность по-разному меняется во время задержки в зависимости от величины первого стимула и от того, какое решение принимает обезьяна: считать частоту второго стимула большей или меньшей по сравнению с частотой первого стимула. Способность нейронов по-разному реагировать на различные стимулы называется селективностью, которая тесно связана с понятием рецептивных полей. Рецептивным полем нейрона обычно называют область в пространстве стимулов, стимуляция которой приводит к выраженному отклику нейрона в виде генерации серии спайков. Другими словами, оно характеризует предпочитаемый данным нейроном стимул, то есть стимул, вызывающий наибольший отклик. Самый известный пример анализа рецептивных полей был проведен в работе нобелевских лауреатов Д. Хьюбела и Т. Визеля, которые обнаружили, что рецептивные поля нейронов зрительной коры имеют пространственно локализованную структуру. Например, есть клетки, которые предпочтительно реагируют на вертикальные полосы зрительного стимула, и если менять угол наклона полос, активность таких нейронов падает. Таким образом, частота спайковой активности является вполне определенной функцией угла наклона зрительного стимула с выраженным максимумом. В рассматриваемой задаче сравнения двух вибротактильных стимулов с задержкой наблюдается *смешанная* селективность, то есть активность нейронов префронтальной коры сложным образом зависит от параметров предъявляемых стимулов и принимаемого решения. Выяснение, каким образом параметры задачи кодируются в популяционной активности нейронов, остается актуальной и неразрешенной задачей нейронауки по сей день.

2. Рекуррентная нейронная сеть: структура, цели и метод обучения

Для моделирования описанной когнитивной функции рассмотрим базовую схему, изображенную на рис. 2. Центральная часть – это рекуррентная нейронная сеть, состоящая из N элементов, активность которых описывается системой в векторной форме:

$$\tau \frac{d\mathbf{r}(t)}{dt} = -\mathbf{r}(t) + \mathbf{f}(\mathbf{W}\mathbf{r}(t) + \mathbf{U}\mathbf{x}(t) + \mathbf{b}_r), \quad (1)$$

где вектор-столбец $\mathbf{r} = [r_j]$, $j = 1, \dots, N$, характеризует активность нейронов в момент времени t , матрица $\mathbf{W} \in \mathbb{R}^{N \times N}$ задает рекуррентные веса сети, то есть коэффициенты межнейронных связей. Нелинейная вектор-функция активации $\mathbf{f}(\cdot)$ задает форму отклика отдельных нейронов в ответ на действие входов и остальных нейронов рекуррентной сети. Сеть имеет два входа, активность которых определяется вектором $\mathbf{x} = [x_1, x_2]^T$, а матрица $\mathbf{U} \in \mathbb{R}^{N \times 2}$ задает веса связей этих входов с нейронами рекуррентной нейронной сети. Также имеется три выхода, активность которых описывается вектором $\mathbf{z} = [z_1, z_2, z_3]^T$, считывающим динамику рекуррентной сети через веса выходной матрицы $\mathbf{V} \in \mathbb{R}^{3 \times N}$:

$$\mathbf{z}(t) = \mathbf{V}\mathbf{r} + \mathbf{b}_z. \quad (2)$$

Слагаемые $\mathbf{b}_r \in \mathbb{R}^{N \times 1}$ и $\mathbf{b}_z \in \mathbb{R}^{3 \times 1}$ – это векторы смещения для нейронов рекуррентной нейронной сети и выходного слоя. Форма записи функции активации $\mathbf{f}(\mathbf{x})$ векторного аргумента

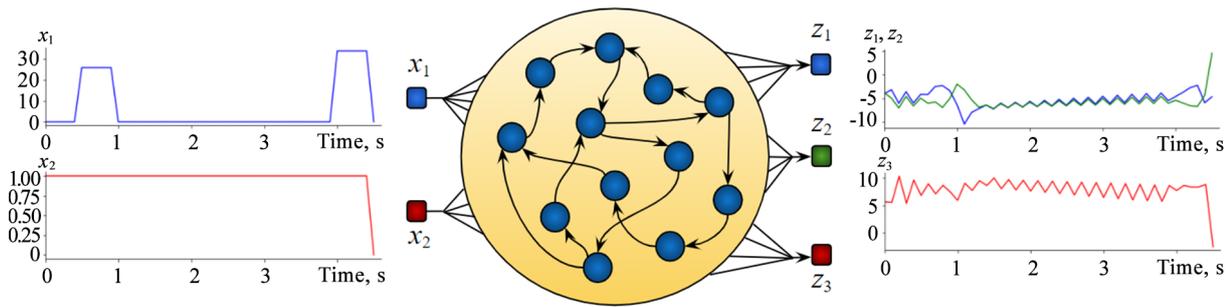


Рис. 2. В центре: схема модели, состоящей из рекуррентной нейронной сети, двух входов x_1, x_2 и трёх выходов z_1, z_2, z_3 . Слева показаны сигналы, подающиеся на соответствующие входы в течение одного испытания. Справа приведены отклики обученной сети с соответствующих выходов

Fig. 2. In the center: scheme of the model consisting of recurrent neural network, two inputs x_1, x_2 and three outputs z_1, z_2, z_3 . On the left the signals applied to appropriate inputs during one trial are shown. On the right the responses are shown of the trained network from the corresponding outputs

$\mathbf{x} = [x_j]$ означает, что компоненты функции вычисляются как $f(x_j)$, то есть $\mathbf{f}(\mathbf{x}) = [f(x_j)]$, $j = 1, \dots, N$.

Обучение рекуррентной нейронной сети (1) происходит в ходе серии испытаний. В процессе каждого испытания нейроны сети получают с первого входа x_1 два стимула f_1 и f_2 , разделенные некоторой задержкой во времени (см. рис. 2). Кроме того, со второго входа x_2 приходит сигнал фиксации, равный 1 в процессе получения обоих стимулов и сбрасываемый до 0 после второго стимула. Цель обучения состоит в том, чтобы при обнулении входного сигнала фиксации $x_2 = 0$ активность выхода, соответствующего большему входному стимулу, была выше другого выхода. То есть при $f_2 > f_1$ на выходе должно быть $z_1 > z_2$ и наоборот. Третий выход – это сигнал фиксации, который изменяется синхронно с входом x_2 .

В общем виде задача контролируемого обучения (обучения с учителем) состоит в настройке параметров $\boldsymbol{\pi}$ искусственной нейронной сети таким образом, чтобы в ответ на некоторый входной стимул $\mathbf{x}^{(i)}$ из множества входов $\{\mathbf{x}^{(i)}\}$, $i = 1, \dots, M$, сеть генерировала выход $\mathbf{z}^{(i)}$, максимально приближенный к целевому сигналу $\mathbf{z}_{\text{targ}}^{(i)}$: $\mathbf{z}^{(i)} = F(\mathbf{x}^{(i)}, \boldsymbol{\pi}) \approx \mathbf{z}_{\text{targ}}^{(i)}$. Другими словами, в процессе обучения минимизируется значение функции ошибки $E = 1/M \sum_i E(\mathbf{z}^{(i)}, \mathbf{z}_{\text{targ}}^{(i)})$, где $E(\mathbf{z}^{(i)}, \mathbf{z}_{\text{targ}}^{(i)})$ – количественная мера отличия целевого сигнала $\mathbf{z}_{\text{targ}}^{(i)}$ от выхода $\mathbf{z}^{(i)}$. Классический метод обучения рекуррентных нейронных сетей основан на стохастическом градиентном спуске. Его суть заключается в рекурсивном изменении параметров $\boldsymbol{\pi}$ в направлении, противоположном градиенту ошибки $\partial E / \partial \boldsymbol{\pi}$. Поскольку оценивать ошибку по всей выборке обучения вычислительно затратно, зачастую ошибка вычисляется по минивыборке – некоторому малому числу $K \ll M$ случайно выбранных элементов $\{\mathbf{x}^{(k)}\}$ с индексами k из набора $\mathbb{S} = \{k_1, \dots, k_K\}$:

$$E_{\text{sample}} = \frac{1}{K} \sum_{k \in \mathbb{S}} E(\mathbf{z}^{(k)}, \mathbf{z}_{\text{targ}}^{(k)}). \quad (3)$$

Для уменьшения ошибки модифицируемые параметры $\boldsymbol{\pi}$ изменяются в направлении, противоположном градиенту, на величину пропорциональную так называемой скорости обучения η :

$$\Delta \boldsymbol{\pi} = -\eta \frac{\partial E}{\partial \boldsymbol{\pi}}. \quad (4)$$

В настоящей работе используется модифицированный метод обучения «Адам» (от «адаптивная оценка моментов») [19], в котором скорость обучения каждого параметра настраивается индивидуально в зависимости от соответствующего среднего значения и стандартного отклонения.

В качестве входных импульсов берутся значения частот, использованных в эксперименте [16], а именно пары величин f_1, f_2 выбираются из списка $\{10, 14, 18, 22, 26, 30, 34\}$. При обучении одна минивыборка состоит из серии предъявлений пар входных стимулов в случайном порядке параллельно с сигналом фиксации в течение определенного отрезка времени (10 с), после чего снятая активность на выходе сравнивается с целевыми значениями и вычисляется ошибка по минивыборке (3). На её основе модифицируются параметры сети (значения W, U, V, b_r, b_z) в соответствии с (4). Этот процесс многократно повторяется до снижения ошибки ниже уровня 0.5%, после чего сеть в 95% случаев даёт правильный ответ. Дальнейший анализ строится для рекуррентной сети из $N = 256$ нейронов с функцией активации типа линейный выпрямитель $f(\cdot) = \text{ReLU}(\cdot)$ ($f(x) = 0$ при $x < 0$ и $f(x) = x$ при $x \geq 0$), скорость обучения $\eta = 10^{-3}$, уравнения (1) дискретизируются с шагом $\Delta t = 0.1$ с. Стоит отметить, что выбранная функция ReLU является одной из распространенных функций активации при построении нейросетей глубокого обучения. В то же время для кортикальных нейронов в физиологически релевантном диапазоне значений параметров эта функция хорошо описывает зависимость частоты спайков от входного тока, отличительной чертой которой является отсутствие значения насыщения [20, 21]. При анализе динамики рекуррентной нейронной сети (1) испытание состоит, как и в эксперименте, из нескольких стадий: в течение 0.5 с подается только сигнал фиксации, затем первый стимул длительностью 0.5 с, после чего задержка 3 с и второй стимул длительностью 0.5 с. После этого сигнал фиксации обнуляется, и выход с максимальной регистрируемой активностью свидетельствует о принятом решении системы о том, какой входной стимул выше.

3. Анализ динамики обученной рекуррентной нейронной сети

После обучения модель качественно воспроизводит поведение обезьян при решении задачи сравнения двух вибротактильных стимулов, предъявляемых с задержкой. Заметим, что при обучении рекуррентная искусственная нейронная сеть ни на какой стадии не «видит» актив-

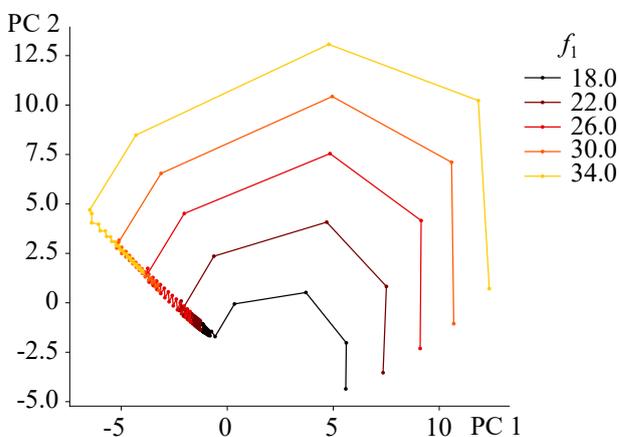


Рис. 3. Траектории на плоскости двух главных компонент, соответствующие активности обученной рекуррентной нейронной сети (1). Различными цветами показаны траектории при различных значениях частоты первого стимула f_1

Fig. 3. Trajectories on the plane of two principal components corresponding to the activity of the trained recurrent neural network (1). Different colors show the trajectories at different values of the frequency of the first stimulus f_1

ности биологических нейронов префронтальной коры, а обучается только на «поведенческих» переменных, которые включают в себя сигналы фиксации, величины двух стимулов и бинарный выходной отклик, соответствующий принятию одного из двух возможных решений. Несмотря на это, оказывается, при выполнении когнитивной задачи модельной сетью многие нейроны во время действия первого стимула и последующей задержки демонстрируют активность, пропорциональную частоте первого стимула, что наблюдается и в биологическом прототипе. В предыдущих исследованиях этой задачи установлено [22], что в основе этого явления лежит формирование множества устойчивых состояний равновесия, кодирующих различные значения стимула f_1 и формирующих рабочую память. На рис. 3 показаны проекции траекторий фазового пространства системы (1) на плоскость двух главных компонент при различных значениях f_1 , где состояния равновесия лежат на некоторой

линии L . Данная плоскость является двумерным срезом N -мерной активности, в ортогональной проекции на который разброс значений $r_i(t)$ максимален. Отстоящие друг от друга отрезки траекторий относятся ко времени действия первого стимула, по окончании которого все траектории медленно движутся вдоль L в течение времени задержки. «Место входа» траекторий в L определяется величиной f_1 .

В настоящей работе исследуются дальнейшие параллели между динамикой искусственной и биологической нейронных сетей при выполнении ими аналогичной функции. Оказывается, что на уровне индивидуальной активности отдельных нейронов при решении рассматриваемой задачи обнаруживаются определенные сходства. На рис. 4 приведены графики активности $r_i(t)$ четырех модельных нейронов после обучения, усредненной по многим реализациям с одним набором параметров. Различными цветами показаны отклики соответствующего нейрона при различных значениях частоты первого стимула f_1 . В обученной сети отдельные нейроны по-разному ведут себя в процессе выполнения задачи. Например, активность нейрона на рис. 4, *a* с приходом первого стимула ($t \approx 0.5$ с) подавляется, а затем резко возрастает, причём тем больше, чем выше частота первого стимула. Во время задержки активность этого нейрона в среднем снижается, но монотонная зависимость от величины первого стимула сохраняется, после чего второй стимул подавляет активность нейрона. Нейрон, активность которого показана на на рис. 4, *b*, с приходом первого стимула понижает свою изначальную активность, причём чем выше величина f_1 , тем сильнее происходит спад. По окончании первого стимула зависимость нейронной активности от f_1 меняется на противоположную: чем выше частота первого стимула, тем больше активность нейрона. После второго стимула его активность снова спадает. В активности нейрона, показанной на рис. 4, *c*, прослеживается следующая закономерность: при стимуляции он активируется,

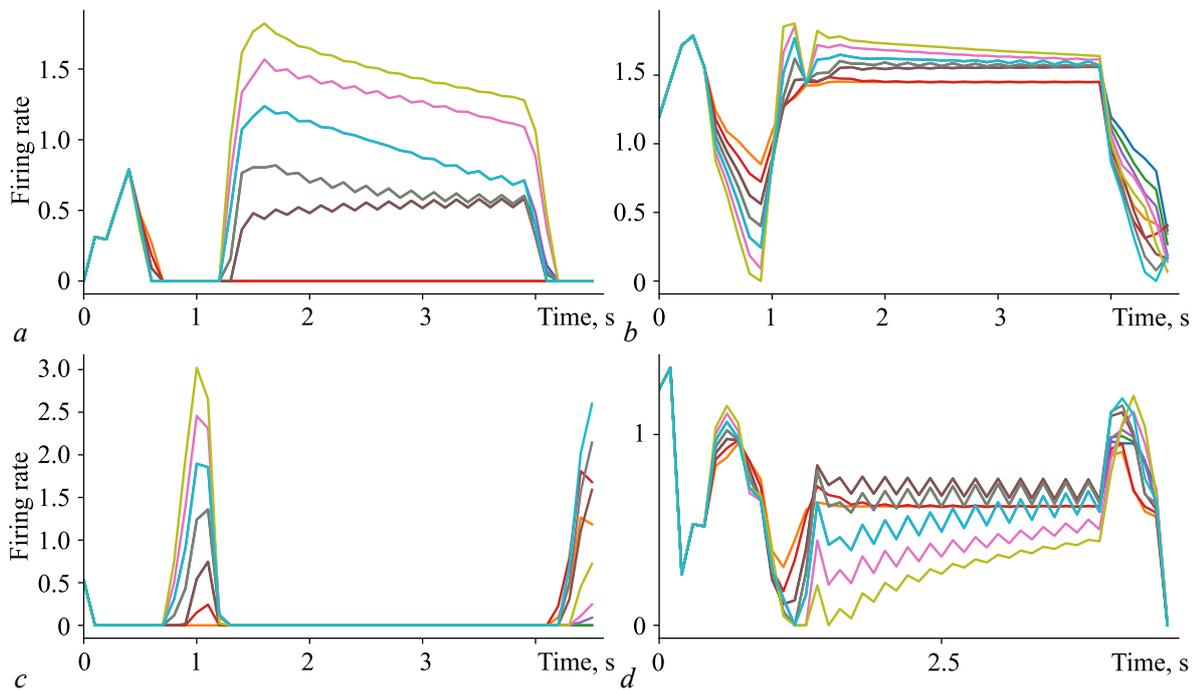


Рис. 4. Активность четырех случайно выбранных модельных нейронов обученной сети (1), усредненная по различным реализациям с одним набором параметров. Различными цветами показана активность данного нейрона при различных значениях частоты первого стимула f_1

Fig. 4. Activity of four randomly selected model neurons of the trained network (1), averaged over different realizations with one set of parameters. Different colors show the activity of a particular neuron at different values of the frequency of the first stimulus f_1

а в отсутствие стимулов его активность стремится к нулю. В продолжение первого стимула наблюдается прямо пропорциональная зависимость степени активности нейрона от частоты стимула. Такая же зависимость во время первого стимула справедлива и для четвертого нейрона на рис. 4, *d*, однако во время задержки степень его активности обратно пропорциональна частоте первого стимула.

Таким образом, при действии стимулов, во время задержки между ними и при принятии решения нейроны сети (1) демонстрируют смешанную селективность, которая присуща и биологическим нейронам. В этом смысле, несмотря на простоту исходных характеристик модели – структуры сети и функции активации нейронов, – можно говорить о некотором сходстве с нейронами префронтальной коры при выполнении искусственной сетью описываемой задачи.

Чтобы определить роль активности нейронов в отображении отдельных аспектов когнитивной задачи, в работе [23] был предложен метод так называемых *разъединенных главных компонент*. В традиционном методе главных компонент находят проекции фазовых траекторий на подпространства с наибольшей дисперсией значений переменных, однако никак не учитывается информация о входах и выходах сети, и, таким образом, результирующие компоненты обладают смешанной селективностью, как и исходные временные ряды активности отдельных нейронов. Суть метода *разъединенных главных компонент* состоит в специальном разложении матрицы данных \mathbf{X} , в которой записана активность нейронов в виде временных рядов средней (мгновенной) частоты генерации спайков. Матрица \mathbf{X} имеет N строк по числу нейронов и T столбцов, где T – длина временных рядов (число дискретов). Также учитывается, что в сеть могут подаваться S различных стимулов, на выходе могут приниматься D возможных решений, и для каждого набора данных параметров проводится K испытаний. Таким образом, i -я строка содержит информацию о динамике средней спайковой частоты i -го нейрона при различных комбинациях входных стимулов и выходных откликов. Разложение матрицы \mathbf{X} представляется в виде:

$$\mathbf{X} = \sum_{\phi} \mathbf{X}_{\phi} + \mathbf{X}_{\text{noise}}, \quad (5)$$

где слагаемые \mathbf{X}_{ϕ} – *разъединенные компоненты* популяционной активности, каждая из которых зависит либо только от стимула (\mathbf{X}_s), либо только от решения (\mathbf{X}_d), либо от их комбинаций, либо не зависит ни от каких условий, а только от времени (\mathbf{X}_t); $\mathbf{X}_{\text{noise}}$ – стохастическое слагаемое.

Несмотря на то, что в активности отдельных нейронов наблюдается широкое разнообразие откликов при действии различных стимулов и при принятии различных решений, после разложения N -мерной активности на *разъединенные компоненты* наблюдаются траектории, форма которых либо не зависит от параметров задачи, либо зависит от величины стимула, либо зависит от одного из возможных принятых решений [23]. На языке фазового пространства это означает, что многомерная активность сети проецируется на некоторое низкоразмерное подпространство, оси (*разъединенные компоненты*) которого соответствуют определенному параметру эксперимента. Иными словами, задача состоит в том, чтобы в многомерном фазовом пространстве найти такие оси (не обязательно ортогональные), в проекции на которые можно увидеть взаимосвязь нейронной активности отдельно со стимулами и с принятием решения. В работе [23] показано, что активность нейронов префронтальной коры обезьяны при решении задачи сравнения вибротактильных стимулов с задержкой действительно можно эффективно разложить на такие компоненты. Таким образом, *разъединенные главные компоненты* выделяют в многомерном популяционном отклике существенные черты, которые отражают когнитивную функцию и которые не очевидны при анализе индивидуальных нейронных откликов или традиционных главных компонент.

В настоящей работе описанный метод адаптирован для анализа данных, сгенерированных модельной системой (1). Проведены численные эксперименты при различных значениях вход-

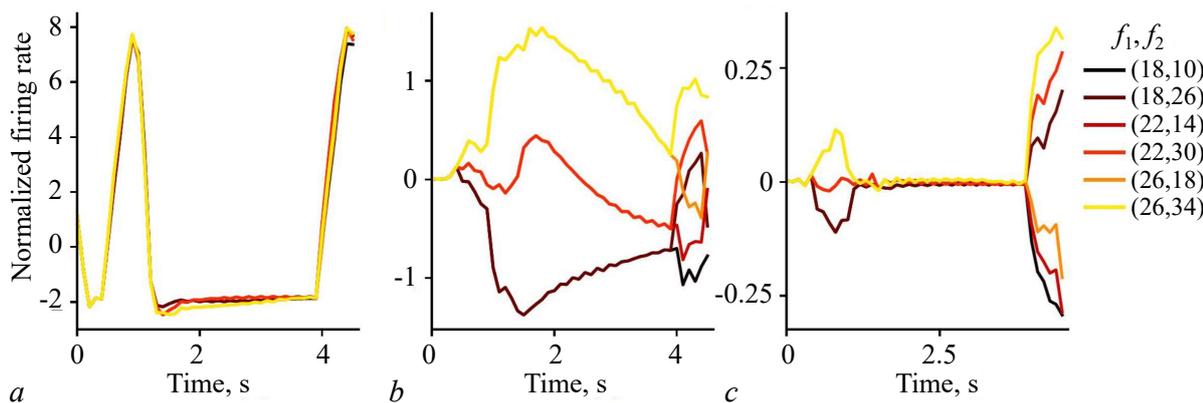


Рис. 5. Разъединенные главные компоненты активности обученной рекуррентной нейронной сети (1) при указанных парах значений f_1, f_2 : *a* – не зависящие от условий, *b* – зависящие от стимула и *c* – зависящие от решения: $f_1 > f_2$ или $f_1 < f_2$

Fig. 5. Demixed principal components of the activity of the trained recurrent neural network (1) for the specified pairs of values f_1, f_2 : *a* – independent of the conditions, *b* – dependent on the stimulus and *c* – dependent on the decision: $f_1 > f_2$ or $f_1 < f_2$

ных стимулов и различных выходных откликах для составления матрицы данных \mathbf{X} , которая разложена на разъединенные компоненты в соответствии с правилом (5). На рис. 5 показаны результаты анализа – разъединенные главные компоненты активности нейронной сети (1) при различных парах значений f_1, f_2 . На рис. 5, *a* показан ход траекторий, которые не зависят от изменения значений стимулов f_1 и f_2 , то есть проекции на данную ось-разъединенную компоненту зависят только от времени. На рис. 5, *b* наблюдается существенное отличие траекторий после первого стимула, то есть соответствующая разъединенная главная компонента, проекции на которую показаны, является стимул-специфичной. На рис. 5, *c* представлены проекции, зависящие от решения $f_1 > f_2$ или $f_1 < f_2$, то есть в данном случае мы имеем дело с разъединенной компонентой, выделяющей в многомерной активности свойства, отражающие принятие решения. Таким образом, метод разъединенных главных компонент, примененный к активности модельной рекуррентной нейронной сети, показал свою эффективность в обнаружении подпространств, динамика на которых отражает отдельные этапы выполняемой когнитивной задачи. Многомерная динамика рекуррентной нейронной сети, моделирующей когнитивную функцию, может быть представлена в более компактном виде в форме низкоразмерных траекторий в специально найденных подпространствах популяционной активности.

Заключение

В работе описан подход построения модели функциональной нейронной сети, активность которой моделирует решение когнитивной задачи. Суть подхода: формулирование целевой функции, задание архитектуры сети, её настройка методами машинного обучения и анализ методами нелинейной динамики. В качестве моделируемого процесса взят экспериментальный прототип – задача вибротактильной дискриминации, которая задействует соматосенсорную рабочую память. В ходе данного эксперимента обезьяна получает два вибростимула с различными частотами, разделенные временной задержкой, после чего нажатием одной из двух доступных кнопок сообщает о том, какой стимул имел большую частоту. Одновременно происходит регистрация активности нейронов префронтальной коры головного мозга. Полученные данные представляют собой многомерные временные ряды гетерогенной нейронной активности, которые сложны для анализа

даже в случае выполнения животным простых задач. Цель данного эксперимента и аналогичных исследований состоит в том, чтобы соотнести параметры когнитивного задания (величину стимулов и выходов) и действий животного с регистрируемой нейронной активностью и сделать соответствующие выводы о функциях мозга.

Для моделирования описанной задачи в настоящей работе предложена рекуррентная нейронная сеть, которая после настройки методами глубокого обучения приобретает форму многомерной динамической системы, состоящей из связанных нелинейных единиц. После поступления входных стимулов с задержкой на выходе сети формируется отклик, информирующий о том, какой стимул больше, то есть система качественно воспроизводит этапы описанного эксперимента. Установлено, что несмотря на простоту исходных характеристик модели – структуры сети и функции активации нейронов, – в многомерной активности модельных нейронов присутствуют черты, имеющие сходство с нейронами префронтальной коры при выполнении искусственной сетью описываемой задачи. Во-первых, показано, что в искусственной нейронной сети наблюдается явление смешанной селективности, характерное для популяции нейронов префронтальной коры. Во-вторых, установлено, что в активности модельной нейронной сети можно выявить разьединенные главные компоненты – «строительные блоки» полной многомерной активности, которые по отдельности устанавливают соответствие динамики нейронов отдельным параметрам (входных стимулов и выходных откликов) выполняемой задачи. Одна из возможных причин возникновения данных аналогий состоит в том, что выбранная функция активации, несмотря на свою простоту, всё же отражает одно из ключевых динамических свойств кортикальных нейронов, для которых характерна близкая к кусочно-линейной зависимость средней частоты спайков от входного тока и отсутствие насыщения [20, 21]. Заметим, что априори нет веских причин, по которым активность нейронов, можно разложить на отдельные низкоразмерные компоненты, соответствующие различным параметрам задачи [24]. Напротив, можно было бы ожидать, что отражение этих параметров в откликах нейронной популяции должно быть существенно высокоразмерным. Таким образом, разложение многомерной активности на разьединенные главные компоненты является нетривиальным результатом, и возникает вопрос о причинах и механизмах этого явления, однако это выходит за рамки данного исследования и остается предметом дальнейшего изучения. В целом, описанный подход показал свою эффективность даже при малом числе исходных предположений о структуре и динамике нейронной сети.

Список литературы

1. *Дмитричев А. С., Касаткин Д. В., Клиньшов В. В., Кириллов С. Ю., Масленников О. В., Щапин Д. С., Некоркин В. И.* Нелинейные динамические модели нейронов: обзор // Известия вузов. ПНД. 2018. Т. 26, № 4. С. 5–58. DOI: 10.18500/0869-6632-2018-26-4-5-58.
2. *Некоркин В. И.* Нелинейные колебания и волны в нейродинамике // УФН. 2008. Т. 178, № 3. С. 313–323. DOI: 10.3367/UFNr.0178.200803g.0313.
3. *Ehrlich D. B., Stone J. T., Brandfonbrener D., Atanasov A., Murray J. D.* PsychRNN: An accessible and flexible python package for training recurrent neural network models on cognitive tasks // eNeuro. 2021. Vol. 8, no. 1. P. ENEURO.0427–20.2020. DOI: 10.1523/ENEURO.0427-20.2020.
4. *Richards B. A., Lillicrap T. P., Beaudoin P., Bengio Y., Bogacz R., Christensen A., Clopath C., Costa R. P., de Berker A., Ganguli S., Gillon C. J., Hafner D., Kepecs A., Kriegeskorte N., Latham P., Lindsay G. W., Miller K. D., Naud R., Pack C. C., Poirazi P., Roelfsema P., Sacramento J., Saxe A., Scellier B., Schapiro A. C., Senn W., Wayne G., Yamins D., Zenke F., Zylberberg J., Therien D., Kording K. P.* A deep learning framework for neuroscience // Nature Neuroscience. 2019. Vol. 22, no. 11. P. 1761–1770. DOI: 10.1038/s41593-019-0520-2.

5. *Barak O.* Recurrent neural networks as versatile tools of neuroscience research // *Current Opinion in Neurobiology*. 2017. Vol. 46. P. 1–6. DOI: 10.1016/j.conb.2017.06.003.
6. *Marblestone A. H., Wayne G., Kording K. P.* Toward an integration of deep learning and neuroscience // *Frontiers in Computational Neuroscience*. 2016. Vol. 10. P. 94. DOI: 10.3389/fncom.2016.00094.
7. *Sussillo D.* Neural circuits as computational dynamical systems // *Current Opinion in Neurobiology*. 2014. Vol. 25. P. 156–163. DOI: 10.1016/j.conb.2014.01.008.
8. *Масленников О. В., Пугавко М. М., Щапин Д. С., Некоркин В. И.* Нелинейная динамика и машинное обучение рекуррентных спайковых нейронных сетей // *УФН*. принята к публикации. DOI: 10.3367/UFNr.2021.08.039042.
9. *Schmidhuber J.* Deep learning in neural networks: An overview // *Neural Networks*. 2015. Vol. 61. P. 85–117. DOI: 10.1016/j.neunet.2014.09.003.
10. *LeCun Y., Bengio Y., Hinton G.* Deep learning // *Nature*. 2015. Vol. 521, no. 7553. P. 436–444. DOI: 10.1038/nature14539.
11. *Maslennikov O. V., Nekorkin V. I.* Stimulus-induced sequential activity in supervisedly trained recurrent networks of firing rate neurons // *Nonlinear Dynamics*. 2020. Vol. 101, no. 2. P. 1093–1103. DOI: 10.1007/s11071-020-05787-0.
12. *Pugavko M. M., Maslennikov O. V., Nekorkin V. I.* Dynamics of spiking map-based neural networks in problems of supervised learning // *Communications in Nonlinear Science and Numerical Simulation*. 2020. Vol. 90. P. 105399. DOI: 10.1016/j.cnsns.2020.105399.
13. *Пугавко М. М., Масленников О. В., Некоркин В. И.* Динамика сети дискретных модельных нейронов при контролируемом обучении системы резервуарных вычислений // *Известия вузов. ПНД*. 2020. Т. 28, № 1. С. 77–89. DOI: 10.18500/0869-6632-2020-28-1-77-89.
14. *Maslennikov O. V., Nekorkin V. I.* Collective dynamics of rate neurons for supervised learning in a reservoir computing system // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2019. Vol. 29, no. 10. P. 103126. DOI: 10.1063/1.5119895.
15. *Масленников О. В., Некоркин В. И.* Адаптивные динамические сети // *УФН*. 2017. Т. 187, № 7. С. 745–756. DOI: 10.3367/UFNr.2016.10.037902.
16. *Romo R., Brody C. D., Hernández A., Lemus L.* Neuronal correlates of parametric working memory in the prefrontal cortex // *Nature*. 1999. Vol. 399, no. 6735. P. 470–473. DOI: 10.1038/20939.
17. *Romo R., Salinas E.* Flutter Discrimination: neural codes, perception, memory and decision making // *Nature Reviews Neuroscience*. 2003. Vol. 4, no. 3. P. 203–218. DOI: 10.1038/nrn1058.
18. *Barak O., Tsodyks M., Romo R.* Neuronal population coding of parametric working memory // *Journal of Neuroscience*. 2010. Vol. 30, no. 28. P. 9424–9430. DOI: 10.1523/JNEUROSCI.1875-10.2010.
19. *Kingma D. P., Ba J.* Adam: A method for stochastic optimization // *arXiv preprint arXiv:1412.6980*. 2014.
20. *Koppe G., Toutounji H., Kirsch P., Lis S., Durstewitz D.* Identifying nonlinear dynamical systems via generative recurrent neural networks with applications to fMRI // *PLoS Computational Biology*. 2019. Vol. 15, no. 8. P. e1007263. DOI: 10.1371/journal.pcbi.1007263.
21. *Hertäg L., Durstewitz D., Brunel N.* Analytical approximations of the firing rate of an adaptive exponential integrate-and-fire neuron in the presence of synaptic noise // *Frontiers in Computational Neuroscience*. 2014. Vol. 8. P. 116. DOI: 10.3389/fncom.2014.00116.
22. *Yang G. R., Wang X.-J.* Artificial neural networks for neuroscientists: A primer // *Neuron*. 2020. Vol. 107, no. 6. P. 1048–1070. DOI: 10.1016/j.neuron.2020.09.005.
23. *Kobak D., Brendel W., Constantinidis C., Feierstein C. E., Kepecs A., Mainen Z. F., Qi X.-L., Romo R., Uchida N., Machens C. K.* Demixed principal component analysis of neural population data // *eLife*. 2016. Vol. 5. P. e10989. DOI: 10.7554/eLife.10989.

24. *Keemink S. W., Machens C. K.* Decoding and encoding (de)mixed population responses // *Current Opinion in Neurobiology*. 2019. Vol. 58. P. 112–121. DOI: 10.1016/j.conb.2019.09.004.

References

1. Dmitrichev AS, Kasatkin DV, Klinshov VV, Kirillov SY, Maslennikov OV, Shchapin DS, Nekorkin VI. Nonlinear dynamical models of neurons: Review. *Izvestiya VUZ. Applied Nonlinear Dynamics*. 2018;26(4):5–58 (in Russian). DOI:10.18500/0869-6632-2018-26-4-5-58.
2. Nekorkin VI. Nonlinear oscillations and waves in neurodynamics. *Phys. Usp.* 2008;51(3): 295–304. DOI:10.1070/PU2008v051n03ABEH006493.
3. Ehrlich DB, Stone JT, Brandfonbrener D, Atanasov A, Murray JD. PsychRNN: An accessible and flexible python package for training recurrent neural network models on cognitive tasks. *eNeuro*. 2021;8(1):ENEURO.0427–20.2020. DOI:10.1523/ENEURO.0427-20.2020.
4. Richards BA, Lillicrap TP, Beaudoin P, Bengio Y, Bogacz R, Christensen A, Clopath C, Costa RP, de Berker A, Ganguli S, Gillon CJ, Hafner D, Kepecs A, Kriegeskorte N, Latham P, Lindsay G, Miller K, Naud R, Pack CC, Poirazi P, Roelfsema P, Sacramento J, Saxe A, Scellier B, Schapiro AC, Senn W, Wayne G, Yamins D, Zenke F, Zylberberg J, Therien D, Kording KP. A deep learning framework for neuroscience. *Nature Neuroscience*. 2019;22(11):1761–1770. DOI:10.1038/s41593-019-0520-2.
5. Barak O. Recurrent neural networks as versatile tools of neuroscience research. *Current Opinion in Neurobiology*. 2017;46:1–6. DOI:10.1016/j.conb.2017.06.003.
6. Marblestone AH, Wayne G, Kording KP. Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience*. 2016;10:94. DOI:10.3389/fncom.2016.00094.
7. Sussillo D. Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology*. 2014;25:156–163. DOI:10.1016/j.conb.2014.01.008.
8. Maslennikov OV, Pugavko MM, Shchapin DS, Nekorkin VI. Nonlinear dynamics and machine learning of recurrent spiking neural networks. *Phys. Usp.* Accepted. DOI:10.3367/UFNe.2021.08.039042.
9. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks*. 2015;61: 85–117. DOI:10.1016/j.neunet.2014.09.003.
10. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444. DOI:10.1038/nature14539.
11. Maslennikov OV, Nekorkin VI. Stimulus-induced sequential activity in supervisory trained recurrent networks of firing rate neurons. *Nonlinear Dynamics*. 2020;101(2):1093–1103. DOI:10.1007/s11071-020-05787-0.
12. Pugavko MM, Maslennikov OV, Nekorkin V. Dynamics of spiking map-based neural networks in problems of supervised learning. *Communications in Nonlinear Science and Numerical Simulation*. 2020;90:105399. DOI:10.1016/j.cnsns.2020.105399.
13. Pugavko MM, Maslennikov OV, Nekorkin VI. Dynamics of a network of map-based model neurons for supervised learning of a reservoir computing system. *Izvestiya VUZ. Applied Nonlinear Dynamics*. 2020;28(1):77–89 (in Russian). DOI:10.18500/0869-6632-2020-28-1-77-89.
14. Maslennikov OV, Nekorkin VI. Collective dynamics of rate neurons for supervised learning in a reservoir computing system. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2019;29(10):103126. DOI:10.1063/1.5119895.
15. Maslennikov OV, Nekorkin VI. Adaptive dynamical networks. *Phys. Usp.* 2017;60(7):694–704. DOI:10.3367/UFNe.2016.10.037902.
16. Romo R, Brody CD, Hernández A, Lemus L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*. 1999;399(6735):470–473. DOI:10.1038/20939.

17. Romo R, Salinas E. Flutter Discrimination: neural codes, perception, memory and decision making. *Nature Reviews Neuroscience*. 2003;4(3):203–218. DOI:10.1038/nrn1058.
18. Barak O, Tsodyks M, Romo R. Neuronal population coding of parametric working memory. *Journal of Neuroscience*. 2010;30(28):9424–9430. DOI:10.1523/JNEUROSCI.1875-10.2010.
19. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014.
20. Koppe G, Toutounji H, Kirsch P, Lis S, Durstewitz D. Identifying nonlinear dynamical systems via generative recurrent neural networks with applications to fMRI. *PLoS Computational Biology*. 2019;15(8):e1007263. DOI:10.1371/journal.pcbi.1007263.
21. Hertäg L, Durstewitz D, Brunel N. Analytical approximations of the firing rate of an adaptive exponential integrate-and-fire neuron in the presence of synaptic noise. *Frontiers in Computational Neuroscience*. 2014;8:116. DOI:10.3389/fncom.2014.00116.
22. Yang GR, Wang XJ. Artificial neural networks for neuroscientists: A primer. *Neuron*. 2020;107(6):1048–1070. DOI:10.1016/j.neuron.2020.09.005.
23. Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Qi XL, Romo R, Uchida N, Machens CK. Demixed principal component analysis of neural population data. *eLife*. 2016;5:e10989. DOI:10.7554/eLife.10989.
24. Keemink SW, Machens CK. Decoding and encoding (de)mixed population responses. *Current Opinion in Neurobiology*. 2019;58:112–121. DOI:10.1016/j.conb.2019.09.004.



Масленников Олег Владимирович – родился в Горьком (1988), окончил радиофизический факультет Нижегородского государственного университета им. Н. И. Лобачевского (2011). С 2007 года работает в Институте прикладной физики РАН. Защитил диссертацию на соискание учёной степени кандидата физико-математических наук (2014, ИПФ РАН). Область научных интересов – нелинейные колебания и волны, синхронизация, сложные сети, нейродинамика. Имеет более 40 научных публикаций по указанным направлениям. Награжден медалью РАН с премией для молодых ученых (2016).

Институт прикладной физики РАН
 Россия, 603950 Нижний Новгород, Ульянова, 46
 E-mail: olmaov@ipfran.ru
 ORCID: 0000-0002-8909-321X
 AuthorID: 643949